Earthmover resilience & testing in ordered structures

Omri Ben-Eliezer
Tel-Aviv University

Eldar Fischer
Technion

Computational Complexity Conference 2018
UCSD, San Diago
Property testing (RS96, GGR98)

**Meta problem:** Given property $P$, efficiently distinguish between

- Objects that **satisfy** $P$
- Objects that are **far from satisfying** $P$
Definition: An $\varepsilon$-test for a property $P$ is given query access to an unknown graph $G$ on $n$ vertices, and acts as follows.

- $G$ satisfies $P$: ACCEPT (with prob. $2/3$)
- $G$ is $\varepsilon$-far from $P$: REJECT (with prob. $2/3$)
- DON'T CARE
Graph property testing

Query = “Is there an edge between u and v?”
(dense graph model)

\[ G \text{ satisfies } P \]

ACCEPT (with prob. 2/3)

\[ G \text{ is } \varepsilon\text{-far from } P \]

REJECT (with prob. 2/3)

DON’T CARE
Graph property testing

$\varepsilon$-far = need to add/remove $\varepsilon n^2$ edges in $G$ to satisfy $P$. (dense graph model)

$G$ satisfies $P$

ACCEPT (with prob. 2/3)  DON’T CARE  REJECT (with prob. 2/3)

$G$ is $\varepsilon$-far from $P$
Graph property testing

**Definition:** A property $P$ is **testable** if it has an $\varepsilon$-test making $q(\varepsilon)$ queries for any $\varepsilon > 0$.

**Question (GGR98):** Which graph properties are testable?
Canonical tests

• An $\epsilon$-test is **canonical** if it queries a random induced subgraph and accepts/rejects only based on queried subgraph.
Canonical tests

• **Theorem** [AFKS00, GT03]:
  \[ P \text{ testable} \iff P \text{ canonically testable} \]

**Intuition:** *Original* test makes \( q \) queries
*Canonical* test picks random \( 2q \) vertices, then “simulates” original test
Tolerant testing [PRR’06]

• Test is \((\delta, \varepsilon)\)-tolerant \((0 \leq \delta < \varepsilon)\) if it acts as follows.
  • Motivation: Noisy input

\[ G \text{ satisfies } P \]
\[ G \text{ is } \delta \text{–close to } P \]
\[ G \text{ is } \varepsilon \text{–far from } P \]

- ACCEPT (with prob. 2/3)
- DON’T CARE
- REJECT (with prob. 2/3)
Tolerant testing [PRR’06]

P is tolerantly testable

∀ε ∃δ : P has a (δ, ε)-test making q(ε) queries.

- G satisfies P
- G is δ-close to P

- G is ε-far from P

ACCEPT (with prob. 2/3)
DON’T CARE
REJECT (with prob. 2/3)
Distance estimation

∀ε ∀δ : P has a (ε − δ, ε)-test making q(δ, ε) queries.

G is (ε − δ)-close to P

ACCEPT (with prob. 2/3)

G is ε-far from P

REJECT (with prob. 2/3)

DON’T CARE
Testing vs tolerant testing vs distance estimation

- **Theorem** [Fischer, Newman ’05]: For graph properties, P canonically testable P estimable

\[
G \text{ satisfies } P \quad \text{ACCEPT} \\
G \text{ is } \varepsilon\text{-far from } P \quad \text{DON'T CARE} \\
G \text{ is } (\varepsilon - \delta)\text{-close to } P \quad \text{ACCEPT} \\
G \text{ is } \varepsilon\text{-far from } P \quad \text{REJECT}
\]
Summary - graph properties

**Testability**
- $G$ satisfies $P$ → ACCEPT
- $G$ is $\varepsilon$-far from $P$ → REJECT
- DON’T CARE

**Canonical Testability**
- $G$ satisfies $P$ → ACCEPT
- $G$ is $\varepsilon$-far from $P$ → REJECT
- DON’T CARE

**Estimability**
- $G (\varepsilon - \delta)$-close to $P$ → ACCEPT
- $G$ is $\varepsilon$-far from $P$ → REJECT
- DON’T CARE

**Tolerant Testability**
- $G \delta$-close to $P$ → ACCEPT
- $G$ is $\varepsilon$-far from $P$ → REJECT
- DON’T CARE

(GT03) (FN05)
What about ordered structures?

- **Strings** (1D)
- **Images** (2D) AKA ordered matrices
- **Vertex-ordered graphs** (2D) and hypergraphs
- Hypercube (high-D): a different story...
Image property testing

Unknown $n \times n$ image $I$ over fixed set of pixels $\Sigma$

Query = “What is the color of pixel in location $(i,j)$?”
Image property testing

Unknown $n \times n$ image $I$ over fixed set of pixels $\Sigma$

$\varepsilon$-far = need to modify $\varepsilon n^2$ pixels in $I$ to satisfy $P$

$I$ satisfies $P$  

$\text{ACCEPT}$ (with prob. 2/3)

$I$ is $\varepsilon$-far from $P$

$\text{REJECT}$ (with prob. 2/3)

$\text{DON'T CARE}$
Image property testing

**canonical test** = pick randomly $t$ rows and $t$ columns, query all pixels in intersection.
String property testing

**Query access** to unknown **string** of length $n$ over **fixed** alphabet $\Sigma$.

**canonical test** = pick randomly $t$ elements and query them.
What about ordered structures?

Do similar characterizations hold for ordered structures?

• **No**, testability/estimability $\not\Rightarrow$ canonical testability
  • Example: “not containing three consecutive 1-s” in 0/1 strings.

• **No**, testability $\not\Rightarrow$ tolerant testability. [Fischer, Fortnow ‘05]
  • Properties based on codes & PCPPs.

• **Yes**, for “global enough” properties. [This work]
Earthmover resilience (strings)

**Flip operation:**

**Definition:** Earthmover distance between strings $S$ and $S'$ is

$$d_e(S, S') = \frac{1}{n^2} \cdot \min\{ \text{number of flips to create } S' \text{ from } S, \infty \}$$

**Definition:** Property $P$ is earthmover resilient if $\exists \delta: (0,1) \to (0,1)$ s.t.

- String $S$ satisfies $P$
- String $S'$ satisfies $d_e(S, S') \leq \delta(\varepsilon)$
- String $S'$ is $\varepsilon$-close to $P$
Earthmover resilience (images)

**Flip operation:**

Definition: **Earthmover distance** between images $I$ and $I'$ is

$$d_e(I, I) = \frac{1}{n^2} \cdot \min\{ \text{number of flips to create } I' \text{ from } I , \infty \}$$

Definition: Property $P$ is **earthmover resilient** if $\exists \delta: (0,1) \rightarrow (0,1)$ s.t.

- image $I$ satisfies $P$
- $d_e(I, I') \leq \delta(\varepsilon)$
- Image $I'$ is $\varepsilon$-close to $P$
Which properties are earthmover resilient?

- All unordered graph properties [trivial]
- All **hereditary** properties of strings, images & ordered graphs [AKNS00, ABF17]
- Global visual properties of images
  - Convexity of the 1’s
  - 1’s form a half plane
  - [This work]: In general, all properties with **sparse boundary** between 1’s and 0’s.
Earthmover resilience vs canonical testing

[This work]:

For string properties $P$,

$P$ earthmover resilient $\iff$ $P$ canonically testable

For image and ordered graph properties $P$,

$P$ earthmover resilient + $P$ tolerantly testable $\iff$ $P$ canonically testable
Canonical testing to estimation

[This work]:

For image and ordered graph properties $P$,

- $P$ canonically testable $\Rightarrow$ $P$ (canonically) estimable

Corollary [ABF17 + This work]:

- $P$ hereditary $\Rightarrow$ $P$ (canonically) estimable
ER properties are similar to graph properties

For earthmover resilient properties of images / ordered graphs:

- Tolerant testability
- Canonical testability
- estimability
Warmup proof: ER\[\rightarrow\] canonical testing in binary strings

**ER** => **piecewise canonical testing**

- Consider **Interval partition** of string into sufficiently many parts.

  ![Interval partition](image)

  - In each interval, make **sufficiently many random queries** to estimate number of 0’s and 1’s.
  - Due to ER, this gives good estimate for distance to \( P \):

    \[
    Distance(S, P) \approx \min_{S' \in P} VD(S, S')
    \]

Where **VD(S,S')** denotes average **variation distance** between the distributions of 0’s and 1’s in each interval.
Warmup proof: ER\(\rightarrow\) canonical testing in binary strings

**piecewise canonical testing** => **canonical testing**

- Interval partition can be approximated by
  - Picking sufficiently many random queries.
  - Partitioning them artificially into intervals.

- Consequently, piecewise canonical tests can be simulated by canonical ones.
Bits from the proof: **Szemerédi regularity lemma**

[Szemerédi ‘75]:

**Any** graph has an equipartition of size $C(\varepsilon)$, so that almost all pairs of parts are $\varepsilon$-regular.

Pair is $\varepsilon$-regular if

$$|d - D| \leq \varepsilon$$

for any pair of subsets of size $\geq \varepsilon N$
Bits from the proof: **canonical testing** --> estimation

• **High level idea - unordered case** [Fischer Newman ‘05]
  • **Step 1**: If $P$ is canonically testable, densities of **small induced subgraphs** among graphs **satisfying $P$ different** from those of graphs **far from $P$**.

• **Step 2**: **regular partitions** of graphs **satisfying $P$ differ** from graphs **far from $P$**.
Bits from the proof: **canonical testing** --> **estimation**

- **High level idea - unordered case** [Fischer Newman ‘05]
  - **Step 1**: If P is canonically testable, densities of **small induced subgraphs** among graphs **satisfying P different** from those of graphs **far from P**.
  
  - **Step 2**: **regular partitions** of graphs **satisfying P differ** from graphs **far from P**.

  - **Step 3**: Estimating which regular partitions a graph has - doable with constant number of queries.

  - **Step 4**: distance of **G** from **P** \( \approx \) min distance of a regular partition for **G** from a regular partition for **P**.
Bits from the proof: **canonical testing** --> **estimation**

- **Our observation**
  - Above scheme essentially works for **multipartite** graphs.
  
  - Given ordered graph $G$, take **interval partition** of the vertices, effectively approximating $G$ by a multipartite graph.
Bonus: Regular reducibility

[Alon, Fischer, Newman, Shapira ‘06]:
A graph property $P$ is canonically testable $\iff$ $P$ can be “described” using regular partitions

[This work]:
Same holds for images and ordered graphs.
Towards a limit object?

- Proofs in [ABF’17] and this work rely on interval partitioning.

- A limit object (graphon-like [BCLSSV05; LS08; BCLSV08]) for images and ordered graphs via interval partitioning?
Other open questions

• Testability + earthmover resilience  $\rightarrow$ canonical testability?

• **More efficient** conversions from testability to estimability
  • Hereditary properties in graphs
    [Hoppen, Kohayakawa, Lang, Lefmann, Stagni ‘16 + ‘17]

• The landscape of property testing
  • “Global” properties seem easy to test [AFKS’00, FN’01, ABF’17, this work]
  • Local properties are easy to test [BKR’17, B’18+]
  • Algebraic structure makes it hard to test [FF’05, FPS’17]
  • Other general results?